Clues to Transmission

From source attribution to real-time tracking using pathogen genomes

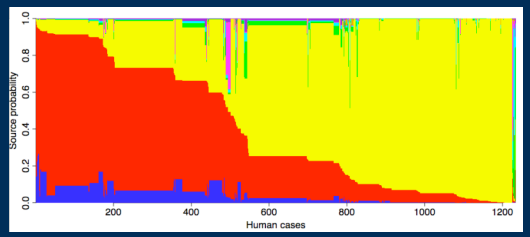
Daniel J. Wilson, Nuffield Department of Clinical Medicine

UNIVERSITY OF OXFORD

The field of Evolutionary Genetics has established itself as an indispensable tool for solving practical problems in Infectious Diseases, providing a framework for understanding the genetic diversity found in pathogen populations [1]. Patterns of genetic variation are informative about transmission, allowing us to reconstruct the epidemiological dynamics of pathogen populations, identify dominant transmission routes, and even to infer direct transmission events between epidemiologically related individuals. As the genetic resolution at which we can feasibly type pathogens has steadily increased, the Evolutionary Genetics revolution has become increasingly ambitious in its goals.

Source Attribution

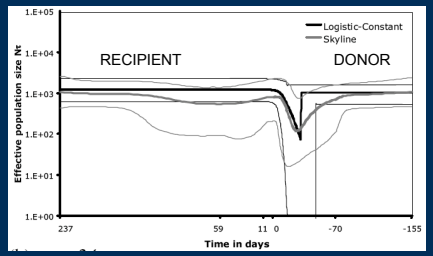
Genomic sequencing is useful for identifying important transmission routes at the population level, an exercise known as *source attribution*. Source attribution models exploit differences in the frequency of alleles between potential source populations to calculate the probability that an isolate of unknown origin arose from a particular population. For example, in a study [2] that



compared 1200 campylobacters isolated from human patients to 1100 animal/environmental campylobacters, pathogenic human bacteria were overwhelmingly found to be genetically typical of bacteria found in animals raised for meat and poultry, implicating food as a major transmission route. The figure illustrates source attribution probabilities for human isolates, principally cattle (red), poultry (yellow) and sheep (blue).

Population Dynamics

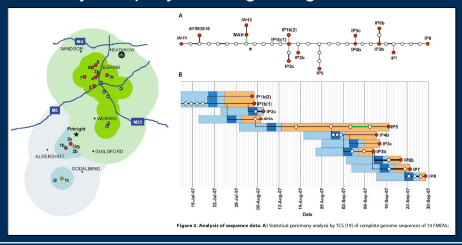
The effect of transmission on population dynamics can be inferred from the signature it leaves on the genome. Bursts of transmission lead to demographic growth in the pathogen population. This expansion is documented in patterns of diversity observable between genomes. When the population size increases quickly, most diversity is likely to have arisen recently, and is therefore at low frequency.



The frequency spectrum of mutations can be used to infer demographic changes that accompany epidemic increases in transmission between hosts, or the establishment of infection within a host. As an example, the figure shows the number of HIV particles that pass the bottleneck during transmission, estimated from genetic diversity within patients [3].

Real-time Tracking

Reconstructing the relatedness of pathogens based on their genomes has enabled the inference of transmission events that are otherwise unobservable, such as the transmission pathways in the 2007 foot-and-mouth disease outbreak (see figure) [4] and direct patient-to-patient transmission in HIV [5]. To date this has been possible only in rapidly evolving viral genomes.



However, now whole-genome sequencing promises to enable the reconstruction of direct transmission events even for slowly-evolving bacterial species. The aim of the *Modernising Medical Microbiology* consortium – a collaboration between the University of Oxford, the Wellcome Trust Sanger Institute, the UK Health Protection Agency and the NHS – is to rapidly detect and track the spread of clinically relevant pathogens in near-to-real time to facilitate control and prevention [6]. Focusing on the pathogens *Staphylococcus aureus*, *Clostridium difficile*, *Mycobacterium tuberculosis* and norovirus, the project will study the nature and extent of hospital and community transmission using genomics to reveal previously unobservable transmission events.

Analytical Challenges

The analysis of whole-genome sequencing at the population level poses many statistical and bioinformatic challenges. Data handling and genome assembly are major obstacles to be overcome before analysis can even begin. Whole-genome data offers unprecedented opportunities to better understand transmission and evolution in pathogens, but only if we can develop the necesary mathematical models and computational tools. Recombination in particular is a prevalent force in bacterial species [7] that considerably complicates statistical analysis. We will need to rely on innovative tools such as ClonalFrame [8] in order to reliably translate patterns of genomic diversity into a proper understanding of microbial transmission.

References

1) Wilson, D. J., D. Falush and G. McVean (2005) Germs, genomes and genealogies. *Trends in Ecology and Evolution* 20: 39-45.
2) Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. Fox, P. Fearnhead, C. A. Hart and P. J. Diggle (2008) Tracing the source of campylobacteriosis. *PLoS Genetics* 4: e1000203.
3) Edwards, C. T. T., E. C. Holmes, D. J. Wilson, R. P. Viscidi, E. J. Abrams, R. E. Phillips and A. J. Drummond (2006) Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC Evolutionary Biology* 6: 28.
4) Cottam, E. M. et al (2008) Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathogens* 4: e1000050.
5) Metzker, M. L., D. P. Mindell, X.-M. Liu, R. G. Ptak, R. A. Gibbs and D. M. Hillis (2002) Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci USA* 99: 14292-14297.
6) Modernising Medical Microbiology website http://www.modmedmicro.ac.uk
7) Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. Fox, C. A. Hart, P. J. Diggle and P. Fearnhead (2009) Rapid evolution and the importance of recombination to the gastro-enteric pathogen *Campylobacter jejuni*. *Molecular Biology and Evolution* 26: 385-297.
8) Didelot, X. and D. Falush (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175: 1251-1266.

EMAIL daniel.wilson@ndm.ox.ac.uk WEB www.danielwilson.me.uk